# THE USE OF TREND CURVES OF RATES OF SPECIES DESCRIPTIONS: EXAMPLES FROM THE CURCULIONIDAE (COLEOPTERA) 

Charles W. O’Brien and Guillermo J. Wibmer

Laboratory of Aquatic Entomology, Florida A \& M University, P. O. Box 111, Tallahassee, FL 32307


#### Abstract

The methods pioneered by Steyskal and White for using trend curves of rates of species descriptions to estimate the total number of extant species in a particular group are reviewed and the most important weaknesses of their methods are discussed in some detail. Trend curves for the Curculionidae for the various biogeographic regions and a cumulative curve for the world are compared. Estimates for the total number of species to be expected are made with each curve and are shown to be invalid. In addition, graphs based on increments of described species by decades are presented for the same regions and the world and these also are analyzed. The results of a statistical approach using a cubic regression program are tabulated. These are shown to be no better than those arrived at through the use of the trend curves. Both estimates (ca. 45,000 to 55,000 species) are demonstrated to be low and are contrasted with the estimate suggested by us $(85,000$ species) based on our combined $30^{+}$year experience in the systematics of Curculionidae.


## Introduction

Steyskal (1965) published the first attempt to estimate the relative current position of taxonomic knowledge in various groups of insects and other animals through the use of trend curves of the rates of species descriptions (represented as cumulative number of species described per decade). White (1975) followed Steyskal's method, and included some additions of his own. He tried to estimate the total numbers of extant species in many of the North American families of Coleoptera, and the year or period when the task of describing them was going to be completed. He purposely excluded 4 large families: Carabidae, Staphylinidae, Curculionidae, and Tenebrionidae, because he assumed they contained appreciable undetected synonymy. Our work arose as an attempt to estimate the final number of species in the family Curculionidae (s.l.) while Dr. J. Howard Frank did the same with the family Staphylinidae. In addition we also wished to compare progress in the rates of species descriptions in Curculionidae in the different biogeographic regions.

Steyskal's method requires obtaining the numbers of species described during specific periods (decades), cumulating totals, plotting them on paper and then joining the dots with a smooth line by using a French curve, trying to join as many dots as possible while attempting to leave half of the remainder on each side of the curve. This means that we trace the curve tending to minimize the sum of the distances (from the points to the curve) in absolute values. As cumulative values were used the curves obtained were always ascending, and since the numbers expected are finite the curves have
an upper asymptote or are tending to it. The curves for all groups in an advanced stage of knowledge were very similarly shaped, sigmoid, and could be divided into 3 sections: (1) basally, asymptotic, nearly horizontal due to a very slow advance, (2) medially, nearly straight, with a very high slope as evidence of a rapid advance, (3) apically, another asymptotic, nearly horizontal portion showing again a very slow advance. The slow rate of descriptions illustrated in section 1 can be explained as a result of early works where knowledge was just being organized. Section 3 was assumed by Steyskal to be an indication that only a small percentage of species in the group remains to be described and named, though he mentioned that "A time of lagging taxonomic work, especially in a small group with few workers, may easily produce such a condition."

White did not include Curculionidae in his work because he assumed there were large numbers of synonymies yet to be discovered, especially of those species described by T. L. Casey. Even though this is probably true, we think that the number of synonymies is probably going to be significant only in the subfamily Baridinae, where Casey worked extensively after 1900, as his work in the Curculionidae prior to that date was in general very good. Also, as we are dealing with the world fauna in a very large family, these numbers will be far less important than in an analysis of the North American fauna only.

## Methods

Initially, our plan was to gather data on description rates of Curculionidae, and then draw some trend curves using the method outlined in Steyskal (1965) and White (1975), which from here on we will call the SteyskalWhite method. Using the weevil sections of the Coleopterorum Catalogus (volumes 27 through 30) we recorded the numbers of species described in each decade from year 0-9 (e.g. 1840-1849, or 1900-1909) with the exception of the first period which only includes 1758-1759. These data were also recorded by biogeographic region in order to facilitate a comparative study of description rates. Each species was recorded once only from the biogeographic region from which it was described. The Coleopterorum Catalogus was used to record data to 1929 and the Zoological Record was used from 1930 to 1969 for the remaining data. For the subfamilies Apioninae and Brachycerinae, the Zoological Record was used from 1910 to 1969 because of the earlier publication date of those fascicles in the Coleopterorum Catalogus. All synonyms were deleted from and all revalidations added to the year in which they were described. This is important if all the data are to be consistent. Hence they reflect the rates of descriptions of species recognized as valid in 1969 as if they had been obtained from an up-to-date catalog.

In general we have followed the traditional limits of the major biogeographic regions, but we were forced to define some of the regions politically in part. This was due to the data recording method of our two primary sources which commonly list localities by country only. We have treated Baja California as Nearctic, but all of the remainder of Mexico is included by us in the Neotropical region despite the fact that most of the species in the northern mountains and deserts have Nearctic affinities. We included all of China and Japan in the Oriental region though the northern areas have many species of Palearctic origin. We also treated New Guinea as Oriental as has become common in insect research.

After our data were obtained, we followed the Steyskal-White method using cumulative values by decades. As was the case with the groups studied before, the curves obtained for the 6 different biogeographic regions and the composite for the world were similarly shaped, though denoting different stages of development (figs. 1-7).

Subsequently we used a cubic regression program prepared by Dr. J. Howard Frank and Mr. G. Alan Curtis to be used in programmable calculators. This program was adapted by Mr. Leonel Zúñiga to be used in a CDC-CYBER-73 computer.

Finally we ran our data in a CDC-CYBER-73 computer using a Statistical Package for the Social Sciences (SPSS), using only the programs "Multiple Regression", "Scattergram" and "Condescriptive". In this case we used both cumulative values and increments.

When correlation coefficients are analyzed, we give the confidence intervals as $(A \pm B[C])$, where $A$ is the mean, $B$ is the standard deviation, and $C$ is the percentage of the mean equivalent to the standard deviation.

## Discussion

Even though our initial plan was to use the Steyskal-White method to make a prediction on the final number of species in the family Curculionidae, we were not going to make a prediction of the year at which that final number would be reached. We believe that there is no possible way to estimate that time because it is theoretically infinite and independent of the total number of species, and will vary depending on the effort made by taxonomists to describe the species. The fact that a small number of species remains to be discovered and named does not mean that it is going to happen in a relatively short period.

This is not even feasible from a practical standpoint since it may lead to a clear contradiction as applied by White. His curve for the families Can-tharidae-Lampyridae-Lycidae indicates an end-point in the years 2090-2100, though the resumé curve (which contains these 3 families) has an end-point in the years 2040-50. The end-point of the resume curve should not occur earlier than the end-point of any individual family considered, but should coincide with the family which will be completed last.

We also disagree with other procedures of the method as follows:

1) Neither Steyskal nor White included dots at those decades where no descriptions were made. If we do this we are not indicating a real lag period, which should have an influence on the shape of the curve. For example, if there are 200 species described up to 1840 and no species are described in the decade 1841-1850, it is logical to use the number obtained in the preceding decade and say that there are 200 species described up to 1850 (which is correct). Otherwise, if a limited number of species is described in the decade (and thus plotted), their influence in decreasing the slope of the curve will be greater than in the case where no species are described (and thus not plotted). This is quite contradictory.
2) In drawing the curves for the Melyridae and Dytiscidae, White purposely connected the curve to the next-to-last dot instead of to the last, as he felt that ". . . interpretation of the significance of the trends on graphs is sometimes essential." This procedure may give a better (more accurate) prediction, but introduces further subjectivity, thereby diminishing the obiective statistical value of the method.

We have illustrated the drastic differences this can produce for the Neotropical region (fig. 2). We feel that the small number of species for the 1960's does not reflect the number of species still to be described, but is the result of numerous variables which are discussed later in this paper. Nevertheless this rather small shift in the curves from the last dot to the next to the last dot produces a difference in the estimated number from a $27 \%$ increase (estimated total A) to a $69 \%$ increase (estimated total B).
3) The result is greatly affected by the time at which the study is carried on. If White's study had been done in 1900, the numbers that would have been predicted for the families Cleridae and Dytiscidae probably would have been about $60 \%$ of those predicted by him in 1975. White already pointed this out using as an example his curve for Cantharidae-Lampyridae - Lycidae. We have no way of correcting this, but it must be taken into consideration as it may be producing biased results.
4) White made an estimation of total numbers of species in certain families of Coleoptera, and also predicted the periods at which those final numbers would be reached. For this he needed to assume that both halves of the curve (and thus what we have called sections 1 and 3 ) are symmetrical. He did not include the reason(s) used to make this assumption. He probably based it on the observation of some nearly complete curves (e.g. birds or butterflies) and then generalized. He used this from a practical standpoint in order to be able to trace the upper half of the curve. In those groups where section 3 was not complete but there was a clear decline in the rate of species descriptions to make possible the estimation of the mid-point (according to his criteria), the bottom half of the curve was transposed to the top in order to make a prediction. There are no bases for assuming this $a$ priori, as the factors determining the development of section 1 are very different from those determining section 3 , and are not necessarily going to produce symmetrical results. A set of data with a symmetrical distribution will produce a sigmoid curve with both halves symmetrical when cumulated. Only 2 of our graphs for the increments (figs. $8 \& 14$ ) denote a tendency to a symmetrical distribution. It is probable that most groups of animals will lack a symmetrical distribution due to the high variability of the factors influencing those data. The selection of the mid-point adds another significant point of subjectivity to this method. We cannot objectively determine the mid-point before the end-point is known. Also, as the total number of species is constant and the year at which they are all going to be known is variable, the mid-point for the number of species is not necessarily going to coincide with the mid-point for the time needed to describe them all.
5) The values obtained are related to only one variable, time, which has the property of varying at a constant rate. The advantage of using time is that it is easy to measure, but there are other variables determining the rate of species descriptions, and some of them may be as much or even more important than the time variable, but unfortunately they are much more difficult to quantify. These factors, separately or together, can produce a lag in the rate of discovery and/or description of new species which can lead to a completely erroneous conclusion. We have grouped what we think are the most important variables into 2 categories, according to whether they can be evaluated quantitatively or qualitatively. Examples of quantitative variables are time, number of active workers, and number of cryptic
species (including siblings). Examples of variables which can be evaluated qualitatively are the number of species described by an individual worker, difficulties associated with the much larger number of species to be known by the taxonomist, the need for much longer, more detailed descriptions in modern taxonomy, the current policy of most publications to require keys and illustrations, the current problem faced by many authors to find publication outlets especially for large papers, the extremely variable rate of species collection in various parts of the world, etc.

Steyskal mentioned that he obtained the best fit between the dots and the curve when the number of workers, etc. did not substantially change but he did not include such factors in the development of his method, probably in order to make it simple. This has the disadvantage that as these factors are highly variable and strongly influence the number of species described, his methods may lead us to wrong results.

All these difficulties induced us to try other approaches to the problem as well. The first method that we used was the cubic regression, and we found that the fit of the calculated data to the observed data was very good. The results obtained with the Steyskal-White method and the cubic regression formula are presented in figs. 1-7 and table 1 respectively. These results did not satisfy us. According to them the level of knowledge of the Curculionidae ( $s . l$. ) is in a very advanced stage, which does not agree with our current concepts of the group.

At present in the collection of one of us (CWO) there are more than 1,000 undescribed species known to us from the Neotropical region and almost certainly many more in groups not as thoroughly studied by us. For the Nearctic region more than 300 undescribed species are on hand in the same collection. In the Neotropics $1 / 3$ to $1 / 2$ of all species being collected by us seem to be undescribed. It seems clear to us that for these 2 regions the amount of work to come is much greater than the estimates show. Indeed the estimate for the final number of species for the Nearctic region (fig. 1) is clearly wrong as already more than 59 species have been described since 1969. Further evidence that this is so are the following examples from some Neotropical systematics papers published in the last 20 years: Kuschel (1957) Epistrophina 34 species known, 62 described as new; Vaurie (1963) Hyphantus 9 species known, 26 described as new; and Howden (1976) Pandeleteius (Colombia and Venezuela only) 11 species known, 53 described as new. This is a common pattern for most papers being published recently on Neotropical weevils.

Our curve for the Australian region deserves special consideration (fig. 6). Using the Steyskal-White or cubic regression methods it would seem very easy to predict the total number of species of Curculionidae in this region, as an asymptote has been reached, and in 1929 and 1939 respectively $95.37 \%$ and $98.76 \%$ of the species currently known had been described. Contrast this estimate with the opinion of Dr. E. C. Zimmerman, who has extensive experience working in that region. He stated (pers. comm.) that "The one [our trend curve] set out for the Australian Region is largely meaningless \& grossly misleading. . . What your curve for Australia indicates is only that Arthur Lea died, \& active descriptive work in the Australian fauna mostly stopped with his death!!!!! It has nothing whatsoever to do with the percentage of the fauna that has been made known... I shall be surprised if much more than $50 \%$ of the fauna of the Australian fauna, in your sense, is de-
scribed. Only a small fraction of the weevils of New Caledonia are described. There are probably several hundred undescribed genera and several thousand undescribed species of weevils in Australia . . . Only a small fraction of the species in some of the subfamilies in Australia has been described. Hence, the curve is totally meaningless in this sense." Dr. Zimmerman also provided us with a paper by Taylor (1976), who estimated, based on the opinions of specialists for the different groups, that only $61 \%$ of the insect species of Australia have been collected and only $45 \%$ are thought to have names available.

Trying to analyze the reasons for the low estimates obtained with the 2 methods used, we observed that there has been a general tendency in all the regions for a drop in the numbers of species described in the last 3 decades, particularly since 1960 . We can explain this as being the result of a change in approach in modern taxonomy, as both taxonomists and editors have become more interested in publishing revisionary works than in mere description of species. This kind of work requires more time on the part of the taxonomist and in general adds comparatively low numbers of newly described species. It tends often to produce larger papers as well and it has become more difficult to find ready publication for very long papers.

We decided to use increments to compare the results with those obtained with cumulative values. The differences found were very striking, and some of them are summarized in tables 2 and 3 , where the correlation coefficients among the different regions and the world are presented. In general, in both tables, the higher values are obtained between the different regions and the world as was expected, as the latter is made up of the other 6 . When the increments are considered (table 2) the correlation coefficients are not too high and fairly spread out ( $0.64492 \pm 0.17240$ [ $26.73 \%$ ]) with a maximum of 0.89168 and a minimum of 0.31263 . Even though the cumulative data do not meet the requirement of independence for the use of correlation coefficients, we decided to present them for comparative purposes (table 3). They are all very high and very close ( $0.98325 \pm 0.01171$ [1.19\%]) with a maximum of 0.99659 and a minimum of 0.96022 . This demonstrates that when cumulative values are used the existing differences among the different regions and the world are greatly reduced. This probably explains why the curves for all the groups studied before were in general similarly shaped, independent of the particular taxonomic history of each one.

If we look at the graphs for the increments (figs. 8-14) we can see that the evolution of the knowledge of the curculionid fauna was not as steady as it seemed to be in the curves for the cumulative data, but there are peaks of different magnitude in the different regions. We observed that these peaks generally correspond to the efforts of 1 to 3 prolific specialists working in the group at that time in that particular area. An excellent example is to be seen in the Neotropical region in the 1900's resulting from Champion's (19021909) phenomenal work in the Biologia Centrali-Americana (fig. 9). We decided then to examine the most significant peaks and include on the appropriate graph the names of those who alone or together were responsible for describing at least $50 \%$ of the species in that period. In most cases the percentage of species described by them was much larger. In all cases the existence or absence of prolific taxonomists in each region at a particular time was more important in the acceleration or diminution in the rate of species description than the percentage of species already known.

Even though we have little data on numbers of undescribed species for the Oriental and Ethiopian regions compared to those we were able to provide for the New World and the Australian region, we can assume that there is a similar situation there, as no major worker has been publishing regularly in the Oriental region since the 1930's and in the Ethiopian region since the mid-1950's.

Year Nearctic Neotropical Palearctic Ethiopian Oriental Australian World


Table 1. Observed $\left(\mathrm{y}_{\mathrm{i}}\right)$ and calculated ( $\hat{\mathrm{y}}_{\mathrm{i}}$ ) cumulative values of the rates of species descriptions in Curculionidae for the different biogeographic regions and totaled for the world. The calculated values were obtained using a cubic regression program and are. presented until they reach a maximum value. Those beyond 1969 are extrapolations. Note that most of these maximums are even lower than the observed data obtained in 1969.

Table 2. Correlation coefficients of the numbers of species of Curculionidae described by decade, among the different biogeographic regions and the world totals.

|  | Nearctic | Neotropical | Palearctic | Ethiopian | Oriental | Australian | World |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Nearctic | 1.00000 |  |  |  |  |  |  |
| Neotropical | 0.33078 | 1.00000 |  |  |  |  |  |
| Palearctic | 0.58931 | 0.48905 | 1.00000 |  |  |  |  |
| Ethiopian | 0.56766 | 0.74370 | 0.49416 | 1.00000 |  |  |  |
| Oriental | 0.77598 | 0.57033 | 0.52783 | 0.85075 | 1.00000 |  |  |
| Australian | 0.76526 | 0.31263 | 0.69271 | 0.40738 | 0.64811 | 1.00000 |  |
| Wor 1d | 0.74132 | 0.81308 | 0.73878 | 0.88097 | 0.89168 | 0.71186 | 1.00000 |

Table 3. Correlation coefficients of the rates of species descriptions of Curculionidae, based on cumulative values by decades, among the different biogeographic regions and the world totals.
Nearctic Neotropical Palearctic Ethiopian Oriental Australian World

| Nearctic | 1.00000 |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Neotropical | 0.97845 | 1.00000 |  |  |  |  |  |
| Palearctic | 0.98788 | 0.96791 | 1.00000 |  |  |  |  |
| Ethiopian | 0.97921 | 0.99650 | 0.96022 | 1.00000 |  |  |  |
| Oriental | 0.98610 | 0.99119 | 0.96140 | 0.99659 | 1.00000 |  |  |
| Australian | 0.99439 | 0.97015 | 0.97616 | 0.97180 | 0.98236 | 1.00000 |  |
| World | 0.99360 | 0.99505 | 0.98206 | 0.99449 | 0.99516 | 0.98762 | 1.00000 |







We believe that the curve obtained for the Palearctic region probably reflects accurately the present knowledge of the group in that area, as the curculionid fauna of the western portion of the region is probably nearly completely known. Probably this is also the case in all regions in groups like birds, mammals or butterflies. Groups which have been traditionally popular among taxonomists and some which contain fairly large and more or less distinct species will commonly produce such curves. It is not reasonable to assume that all animals will fit such curves and this is clearly illustrated by the curves of the curculionids in other regions.



For all the reasons stated earlier we have to say that unfortunately the use of trend curves of rates of species descriptions does not give an accurate prediction of the total number of species extant. This is also the case with the cubic regression method. Table 1 lists the rates of species descriptions in Curculionidae observed, and those calculated using a cubic regression program. The calculated values are presented until they reach a maximum and those beyond 1969 are extrapolations. Since 5 of 7 of these maximums are lower than the observed data obtained in 1969, it is clear that this


method is even poorer than the Steyskal-White method. We did not draw curves for the estimated values because these would be so similar to those presented in figures 1-7 that this would serve no useful purpose. Frank \& Curtis (1979) explain clearly why a logistic equation should be used to describe curves of this nature, as the values obtained with it are always positive and tend to an upper asymptote. We agree with them on this point, but no method is going to generate an accurate prediction because the rate in the number of species descriptions is not only a function of time but of a combination of several variables.



We believe that it is much better to survey the opinions of specialists in the different groups, when they have a broad knowledge of the group and long experience working on it. We feel that the knowledge of the current number of valid species is a great aid in making a decision concerning the total number of extant species. This procedure is not as subjective as it seems to be because it makes use of a very important factor: experience.

The estimate for total numbers to be expected when all species are described as indicated by our world curve, 50,145 species (fig. 7), agrees quite well with Arnett's (1967) estimate of 50,000 . The cumulative estimates from our separate regional curves (figs. 1-6, including curves A and B for the Neotropics) 51,131 and 56,618 are also not too far from this (ca. $2 \%$ and $13 \%$ respectively). However we consider these estimates all to be extremely low. Based upon our combined $30^{+}$years of systematics experience with Curculionidae and a careful study of the evidence in part presented earlier, we believe that there are at least 85,000 species in the world. We believe that in the tropical regions half or less than half of the species of Curculionidae are described. Admittedly, this estimate is a subjective one, but we feel it has greater validity than those produced by the trend curves.

While we consider 85,000 species to be a reasonable number for this family, it is probable that they will never all be described. As Raven (1976) has written, the tropical forests are being totally destroyed at a remarkably rapid rate and with the forests, thousands of species of animals (including insects) are being destroyed as well. Raven states that all of the tropical forests will be destroyed in the next 25 years along with their communities of plants and animals. As many of these tropical communities have not been visited by scientific collectors particularly with modern equipment (e.g., U.V. traps) or with special microhabitats in mind (e.g., litter to be treated in berlese funnels) many species will become extinct before they can be collected and studied.

## Acknowledgments

We wish to thank J. Howard Frank (who is responsible for initiating this study) and his colleague G. Alan Curtis for extensive assistance with the statistics and statistical programs used herein and for critically reading this manuscript and for their numerous suggestions. Our special thanks also go to Leonel Zúñiga, who assisted greatly through his use of computer programs and with his broad knowledge of applied statistics. He and Carlos Pereira critically read the manuscript and made many valuable and useful comments.

This study was supported in part by the following grants in aid: U.S.D.A. Cooperative State Research Service Grant Nos. 716-15-22 and 416-15-16 to Florida A. and M. University.

## Literature Cited

Arnett, R. H., Jr. 1967. Present and Future Systematics of the Coleoptera in North America. Ann. Ent. Soc. Amer. 60(1):162-170.
Champion, G. C. 1902-1909. Biologia Centrali-Americana, Insecta, Coleoptera (Curculionidae: Curculioninae) 4(4):1-729, (5):1-497, (7):1-78.
Frank, J. H. and G. A. Curtis. 1979. Trend lines and the number of species of Staphylinidae. Coleopt. Bull. 33(2):133-149.
Howden, A. T. 1976. Pandeleteius of Venezuela and Colombia (Curculionidae: Brachyderinae: Tanymecini). Mem. Amer. Ent. Inst. (24):1-310.
Kuschel, G. 1957. Revisión de la subtribu Epistrophina. Rev. Chilena Ent. 5:251-364.
O'Brien, C. W. and G. J. Wibmer. 1978. Numbers of genera and species of Curculionidae (Coleoptera). Ent. News 89(2-3):89-92.
Raven, P. H. 1976. The Destruction of the Tropics. Frontiers 40:22-23.
Steyskal, G. C. 1965. Trend Curves of the Rate of Species Descriptions in Zoology. Science 149(3686):880-882.
Taylor, R. W. 1976. Senate Standing Ctee on Science and the Environment. A submission to the inquiry into the impact on the Australian environment of the current woodchip industry programme. Australian Senate Official Hanshard Rep., pp. 3724-3731.
Vaurie, P. 1963. A revision of the South American genus Hyphantus (Coleoptera, Curculionidae, Otiorhynchinae). Bull. Amer. Mus. Nat. Hist. 125(4):239-304.
White, R. E. 1975. Trend curves of the rate of species description for certain North American Coleoptera. Coleopt. Bull. 29(4):281-295.

## List of Figures

Figs. 1-7. Trend curves of rates of species descriptions of Curculionidae, based on cumulative values, by biogeographic region and totaled for the world, produced by using the Steyskal-White method. See O'Brien and Wibmer (1978) for numbers of genera and species recorded by subfamily.

Figs. 8-14. Graphs of numbers of species of Curculionidae described per decade, by biogeographic region and totaled for the world. Names over major peaks indicate authors who alone or together are responsible for $50 \%$ or more of the described species in that decade.

